AD-A043 494 RICE UNIV HOUSTON TEX DEPT OF ELECTRICAL ENGINEERING F/G 6/4
AN ALGORITHM FOR OPTIMAL NONLINEAR STRUCTURE PRESERVING FEATURE--ETC(U)
JUN 77 S A STARKS, R J DE FIGUEIREDO AF-AFOSR-2777-75
UNCLASSIFIED EE-TR-7708 AFOSR-TR-77-1141 NL

| OF |
AD A043494

END
DATE FILMED
9 -77
DDC

(2)

RICE UNIVERSITY

GEORGE R. BROWN SCHOOL OF ENGINEERING

DEPARTMENT OF ELECTRICAL ENGINEERING

HOUSTON, TEXAS

AN ALGORITHM FOR OPTIMAL NONLINEAR STRUCTURE PRESERVING FEATURE EXTRACTION

by

Scott A. Starks* and Rui J. P. de Figueiredo †*

*Department of Electrical Engineering
†Department of Mathematical Sciences

Rice University, Houston, Texas 77001; JUNE 1977

TECHNICAL REPORT EE 7708

Note: This paper is to be presented at the 1977 IEEE International Symposium on Information Theory, to be held at Cornell University, Ithaca, N.Y. on October 10 through 14, 1977.

ABSTRACT:

This paper presents a new approach to nonlinear structure preserving feature extraction. The ideas behind this method were first introduced in (1). This method is based on certain graph theoretical considerations (such as the minimal spanning tree, edge inconsistency, and diameter edges) and topological considerations (such as interpoint distance measures). After introduction of the subject matter and appropriate background material, the algorithm is formulated in section 5.

Numerical results from the application of this algorithm to various test data sets are presented. Evaluation of these test results are quite encouraging.

This work was supported by Air Force Office of Scientific Research under the Grant 75-2777

(See 1473)

1. Introduction

The problem of developing an efficient method of feature extraction is one of the most significant ones in the field of pattern recognition. Due to its importance, feature extraction has recieved much attention in the literature. Good reviews of this subject can be found in (2) and (3).

However most of the methods for feature extraction which exist have been developed in the context of fairly rigidly defined problems. As a result, relatively little work has been done in developing general approaches to feature extraction which are not problem dependent.

The general approaches for dimensionality reduction which have been devised tend to be based upon information theoretic and statistical foundations. In particular many such methods seek to minimize the probability of error or some bound on the probability of error. Such approaches usually assume that a normal probability density function underlies each class. In practice, such an assumption may only be partially true. In other cases, even if the underlying probability density functions were normal, it is virtually impossible to retrieve reliable extimates of the class conditional statistics given a limited number of training samples. Consequently application of general feature extraction techniques based solely on information theoretic or statistical grounds is often inappropriate.

Other general methods of feature extraction based on Karhunen-Loève theory have been developed (4). However these methods require an estimate of the lumped covariance matrix. If one is given only a small number of training samples from a measurement space of high dimensionality, a reliable estimate of this covariance matrix cannot be made. Therefore in this case, feature extraction techniques based upon the Karhunen-Loève expansion are inapplicable.

What is needed is a feature extraction technique which meets the following requirements. First, it should be general in nature, that is, it should not be limited to application on only a certain type of data. Secondly, the feature extraction should be applicable to situations where the ratio of the number of training samples to the dimension of the measurement space is small. As a result, this method should not be hindered by requiring a complete knowledge of the underlying class conditional probability density functions. Also, the feature extraction method should be computationally efficient.

In the present report, we formulate a new approach to nonlinear feature extraction which meets the above requirements. The proposed method is based on optimally preserving certain graph theoretical and topological attributes present in a data set. In the following sections, the mathematical groundwork for this feature extraction method is laid and the attributes which constitute structure

in a data set are studied. The feature extraction algorithm is then formulated and some numerical results obtained from the application of this technique on various data sets are given.

2. Mathematical Groundwork

Feature extraction essentially amounts to finding a transformation from one space called the measurement space to one called the feature space. Observations in the measurement space can be thought of as resulting from the conversion of some physical excitation into raw data by some sensing device or system. The feature extractor then transforms this raw data into a set of variables called features which hopefully provide enough information to allow for correct classification. The block diagram of a typical pattern recognition system is given in Fig. 1.

At this point, we shall state the general problem which this paper addresses. Let there be given a set of N data vectors $X = (X_1, X_2, \ldots X_N)$ where each X_i is a vector which belongs to a real measurement space of dimensionality n. We wish to find a corresponding set of vectors $Y = (Y_1, Y_2, \ldots Y_N)$ where each Y_i is a vector belonging to a real feature space of dimension m. We require that m be an integer satisfying $1 < m < n$. We wish that Y be constructed in such a manner that the structure present in X is optimally preserved under the mapping from X to Y. That is, we wish

that the mapping take place with the least degradation of structure. Obviously, the question arises as to what constitutes structure in a data set. This paper shall address this question later. For the time being, we shall summarize the way in which other authors have approached the general nonlinear feature extraction problem posed earlier.

3. Background

In (5), Sammon developed a nonlinear mapping technique for data structure analysis. In his algorithm, Sammon optimized a criterion functional based on interpoint distance preservation using a steepest descent method. If d_{ij} represents the Euclidean distance between measurement vectors X_i and X_j and d^*_{ij} represents the Euclidean distance between corresponding feature vectors Y_i and Y_j; then the mapping criteria is defined as:

$$Q(Y) = \frac{1}{d}\sum_{i<j}^{N} \frac{(d_{ij} - d^*_{ij})^2}{d_{ij}} \tag{3.1}$$

where $$d = \sum_{i<j}^{N} d_{ij} \tag{3.2}$$

and $X \in R^m$ and $Y \in R^n$.

In general, structure preservation in this approach is maintained by fitting the N points in the feature space such that their interpoint distances best approximate the corresponding interpoint distances in the measurement space. If $Y_i = (Y_{i1}, \ldots Y_{im})^T$, examination of (3.1) shows that the criterion functional is based upon the (N x m) variables y_{ij}, $i=1,\ldots N$; $j=1,\ldots m$.

In certain cases, the product (Nxm) may be so large that optimizing (3.1) would be prohibitive in terms of computational complexity. As a result, Chang and Lee presented a heuristic relaxation method for the solution of the nonlinear feature extraction problem in (6). Other researchers in the area include Shepherd (7), Calvert (8), and White (9).

In all the above cases, the same criterion functional (3.1) is used. In other words, interpoint distance is treated as the only depository of information about a data set. This description of structure in a data set seems somewhat incomplete. It seems that further examination into what constitutes structure in a data set should be made. In the following section we do just this.

4. Structure in a Data Set

When the situation arises that one is confronted with analyzing data from a high-dimensional space under the handicap of having only a small number of observations, he often resorts to cluster analysis to learn something about the structure of the data. Cluster analysis has been studied by a number of authors. For good reviews on this subject, one may refer to (10), (11), and (12).

Let us consider the problem of clustering a set of M objects $Z = (z_1, z_2, \ldots, z_M)$ into C clusters $W_1, W_2, \ldots W_C$. We may think of clustering as no more than the process of

partitioning M objects into C mutually exclusive groups. Similarly, clustering can be thought of as merely a process of assigning one of C membership labels to each of the M objects.

One large category of clustering algorithms is that of hierarchical clustering. Hierarchical clustering consists of a number of different methods which are sequential in nature. Such methods are fully described in (13).

The two major types of hierarchical clustering are agglomerative and divisive hierarchical clustering. In agglomerative hierarchical clustering schemes, one begins by placing each of the M objects into M singleton clusters. At the next stage of the procedure, a new partition is obtained by joining the two "closest" clusters into one cluster, thus diminishing the total number of clusters by one. Here, "closest" depends on how one defines the distance between two clusters. This process is repeated until the desired number of clusters is obtained. Divisive hierarchical clustering is similar except that one begins by placing all M objects into one cluster. Then at each stage of the procedure, one cluster is split into two clusters.

Regardless of the method employed, the way one defines the distance between two clusters is critical. Two of the most commonly used measures of distance between clusters are

$$d_{min}(W, W') = \min_{\alpha \in W, \beta \in W'} d(\alpha, \beta) \qquad (4.1)$$

$$d_{max}(W, W') = \max_{\alpha \in W, \beta \in W} d(\alpha, \beta) \qquad (4.2)$$

where W and W' are clusters and $d(\cdot,\cdot)$ represents some distance measure such as Euclidean distance

Consider the use of d_{min} as a measure of distance between clusters in an agglomerative hierarchical scheme. We may think of the data points as nodes of some graph. A review of graph theory can be found in (14). As a result of using d_{min} nearest neighbors determine the nearest or "closest" subsets or clusters. Suppose that we represent the merger of two clusters W and W' by adding an edge between the nearest pair of nodes, one in W and the other in W'. Since we are always joining distinct clusters, the graph resulting from these edge additions will contain no loops or circuits. Such a graph is termed a tree. If the process of adding edges continues until all nodes have at least one edge connected to them, the resulting graph is termed a spanning tree. It can be proven that that if at each stage of the algorithm, the nearest pair of clusters are merged then the sum of the edge lengths for this spanning tree will never be greater than the sum of the edge lengths of any other spanning tree. Thus this graph is said to be the minimal spanning tree (MST) for the data set and the clustering procedure is called the nearest neighbor or single-linkage algorithm.

When the d_{max} measure of distance is incorporated in an agglomerative scheme, we obtain the so-called furthest neighbor clustering algorithm. This algorithm also has a

graph theoretical analogy. At each stage of the hierarchy, we produce a graph consisting of a number of complete subgraphs representing the clusters. A complete subgraph is one in which each pair of nodes is connected by an edge. Referring to the definition of d_{max} we deduce that the distance between two distinct clusters is determined by the length of the edge connecting the most distant pair of nodes. This quantity is referred to the diameter of the union of the two clusters.

As was seen previously, the minimal spanning tree is intimately connected with the single-linkage clustering algorithm. The MST is a deceptively simple structure which was first introduced by Prim in (15). Zahn in (16) was the first to demonstrate its amazing powers in handling clustering problems which had previously defied solution. He showed that clusters in a two-dimensional space which the eye identified immediately as separate entities could be separated trivially by an algorithm based on the MST.

An important attribute associated with the MST is that of edge inconsistency. The inconsistency of an edge of the MST is defined in the following manner. Suppose that an edge of the MST connects points X' and X''. The inconsistency measure of the edge is defined as the ratio of the length of the edge connecting X' and X'' divided by the average length of all edges connecting either X' or X'', but not both. Zahn states that when the value of edge inconsistency exceeds a

value of 2 then the human eye tends to perceive the components connected to the endpoints of the edge as separate entities. In fact, Zahn states that the MST is the fundamental mechanism explaining proximity and Gestalt effects in psychology.

The main properties of the MST can be summarized as follows:

1. Any node is connected to at least one of its nearest neighbors.

2. Any subtree is connected to at least one of its nearest neighbors by the shortest available path.

3. The MST minimizes all increasing symmetric functions of interpoint distance.

4. The MST connectivity is invariant under any mapping which preserves the rank order of interpoint distances.

5. The MST is easy to compute and it resembles a loopless skeleton of the configuration.

In Fig. 2, the minimal spanning tree for a set of 2-dimensional data points is given.

In the following section, we incorporate the concepts of the MST, cluster diameter, and edge inconsistency into a method for nonlinear feature extraction.

5. Feature Extraction Procedure

For convenience, we shall now restate the feature extraction problem:

Let there be given a set of points $X = (X_1, X_2, \ldots X_N)$ where each $X_i \in R^n$. We wish to find a corresponding set of points $Y = (Y_1, Y_2, \ldots Y_N)$ where each $Y \in R^m$ $(1<m<n)$ in such a manner that the structure contained in the data set X is optimally preserved under the transformation $g: R^n \rightarrow R^m$ which maps X into Y. Due to the one-to-one correspondence between sets X and Y, we take the liberty of expressing $g(X_i)$ as Y_i.

We begin the feature extraction procedure by computing the $N(N-1)/2$ independent interpoint distances for the set X. Once this is done, the minimal spanning tree for the data set can be constructed. After this is done, we compute the edge inconsistency measure for each of the $(N-1)$ edges of the tree. These values for edge inconsistency are stored in an array E of dimension $(N-1)$.

In addition to array E, we introduce the array C which shall be used in the feature extraction process. Since we may think of clustering as merely a method of labelling points according to cluster membership, we can store such membership labels in the array C. The i^{th} component of C, denoted C_i, will contain the cluster membership of point X_i. Initially, we place all points of X into cluster #1 so $C_i = 1$ for all values of i initially.

Another array B of dimension N is utilized in the procedure. Since the feature extraction process proposed is a sequential one, it is necessary to keep track of which vectors have been transformed. We use the array B to do

this. We initially set all the components B_i of B equal to 0, to reflect the fact that none of the vectors have been transformed. As we transform a vector, say X_i, we set B_i to 1 and it remains at that value for the remainder of the procedure.

To begin the feature extraction process, we search the array E to find the MST edge with the largest value of edge inconsistency. The endpoints of this edge are determined and are denoted by X_k^1 and X_l^1. At this time, the diameter edge is found and the endpoints of this edge are denoted by X_m^1 and X_n^1. Once the identity of these four points is known, we group them into what is called the active set, A^1. This set, $A^1 = (X_k^1, X_l^1, X_m^1, X_n^1)$ is termed the active set because it contains the vectors whose images Y_k^1, Y_l^1, Y_m^1, and Y_n^1 are to be located optimally in the feature space. The superscript indicates that we are at the first stage of the feature extraction process.

Once the active set membership is determined, we update the array B by setting B_i to 1 for Y_i in A^1. We then proceed to find the feature or image space configuration $g(A^1) = (Y_k^1, Y_l^1, Y_m^1, Y_n^1)$ which minimizes the following criterion.

$$Q^1(g(A^1)) = \sum_{i \in I^1} \sum_{j=1}^{N} \frac{(d_{ij} - d_{ij})^2}{d_{ij}} \Delta(1-B_j) \qquad (5.1)$$

where $I^1 = (i : X_i \in A^1)$, (5.2)

d_{ij} is the distance between X_i and X_j, d_{ij} is the distance between Y_i and Y_j, and $\Delta(\cdot)$ is the standard Kronecker delta

function.

The above (5.1) is the standard Sammon criterion except that it is applied to only the four data points of the active set. In addition we do not require that the distance metric be Euclidean. The optimization of (5.1) may be carried out by an appropriate multivariable optimization technique.

Before beginning the next stage of the algorithm, we must update the cluster membership vector C. Suppose that we delete the most inconsistent edge from the MST. The resulting graph would then consist of two connected subgraphs, representing two clusters. We arbitrarily take one of these clusters and relabel it as W_2. To reflect this, we set C_i to 2 for all i such that Y_i is in the selected cluster W_2. Of course, the vectors contained in the other subgraph remain in cluster W_1 and there is no need to update the components of C relating to them.

In general, we repeat the above procedure until all points in X have their images mapped into Y or until the value of edge inconsistency fails to exceed 2. When the value of edge inconsistency fails to exceed 2, then we know that the edge under consideration is not an edge separating what the eye would perceive as two distinct entities or clusters. As a result, we modify our approach. This modification will be discussed later.

To formulate the general approach, suppose that we are

at the K^{th} stage of the algorithm. We conduct a search of the array E to find the K^{th} most inconsistent edge of the MST. Suppose that this edge has a value of inconsistency greater than 2. Furthermore assume that this edge is contained in cluster W_P and that it has endpoints X_k^K and X_l^K. We then find the diameter edge of cluster W_P and determine its endpoints X_m^K and X_n^K. From these four points X_l^K, X_k^K, X_m^K, and X_n^K, we find the points which have not yet had their images located in the feature space and we place these points in the active set A^K. We then update the B vector by setting B_i to 1 for each X_i contained in the active set A^K. We then find the configuration $q(A^K)$ which minimizes the following criterion functional.

$$Q^K(\, g(\, A^K)) = \sum_{i \in I^K} \sum_{j=1}^{N} \frac{(d_{ij} - d_{ij})^2}{d_{ij}} \Delta\, (1-B_j) \qquad (5.3)$$

$$I^K = (\, i \,:\, X_i \in A^K) \qquad (5.4)$$

and d_{ij} and d_{ij} are as defined before.

Once the minimization is performed the cluster membership vector C is updated by the method described earlier and the $(K+1)^{st}$ stage of the algorithm is begun.

If the value of edge inconsistency fails to exceed 2 at at the K^{th} stage of the algorithm, we realize that such an edge does not connect what would readily be perceived as two distinct entities. As a result, we would like to stress intracluster relationships at this stage. To reflect this, we make a modification on the criterion functional expressed in (5.3). Instead of (5.3), we minimize the following

modified criterion functional:

$$Q^K(g(A^K)) = \sum_{i \in I^K} \sum_{j=1}^{N} \frac{(d_{ij}-d_{ij})^2}{d_{ij}} \Delta(1-B_j)\Delta(C_j-P) \quad (5.5)$$

where all quantities I^K, d_{ij}, and d_{ij} are as defined before. The modified criterion functional stresses intracluster over intercluster relationships by virtue of the second Kronecker delta term. Recall that the K^{th} most inconsistent edge fell in cluster W_P. Therefore we concern ourselves only with interpoint distances among members of the active set and members of W_P already located in the image space. In this way we emphasize intracluster structure and at the same time simplify the optimization problem. After (5.5) is minimized we return to the $(K+1)^{st}$ stage of the algorithm and repeat.

A flowchart for the sequential nonlinear structure preserving feature extraction algorithm is presented in Fig.3. One major advantage of this method over other existing methods is that at any stage of the algorithm we are concerned with optimizing a criterion which is dependent on at most (4xm) variables. This greatly reduces the optimizational complexity of the problem.

The algorithm has been programmed in FORTRAN at the Institute for Computer Services and Applications at Rice University. Some numerical results obtained by applying this method to several different data sets are presented in the next section.

6. Numerical Results

The feature extracion algorithm was applied to three different data sets to test its utility. In each case, the dimension of the feature space was chosen to be 2 so that graphical displays of the tranformed data could be obtained. The test data sets were as follows:

1. Artificially generated data: Fifteen points in a 4-dimensional space were generated from three Gaussian pattern classes. Each pattern class shared the same covariance matrix (the identity matrix). The means of the pattern classes were located at the vertices of an isosceles triangle. The resulting transformed configuration is shown in Fig. 4.

2. Nonlinear data: This data set consisted of 30 points distributed along a nonlinear curve in a 5-dimensional space The parametric equations governing this curve were

$$U_1 = \cos U_5 \tag{6.1}$$

$$U_2 = \sin U_5 \tag{6.2}$$

$$U_3 = 0.5 \cos 2U_5 \tag{6.3}$$

$$U_4 = 0.5 \sin 2U_5 \tag{6.4}$$

$$U_5 = .707\ t \tag{6.5}$$

where $t = 0,1,\ldots 29$. This data set was found in 5 . The resulting 2-space configuration is shown in Fig. 5.

3. Ultraviolet fluorescence spectragraphic data: Sixteen ultraviolet fluorescence spectra representing 3 classes of oils (crudes, diesels, and no.6 fuel oils) were sampled at

15 wavelengths. The resulting vectors were mapped into a 2-space as indicated in Fig. 6.

All computer experiments were conducted on an IBM 370/155 general purpose computer. The feature extraction algorithm requires 60 Kbytes of memory. In each case, Euclidean distance was used as the distance metric.

Examination of Fig. 4 reveals that the three Gaussian classes are fairly well separated. Similarly, Fig. 6 indicates a good separation of the three classes of oils. In Fig. 5, we clearly see the string-like structure present in the nonlinear data set. Although in all examples the feature space was 2-dimensional, there is no restriction that this need always be the case.

7. Conclusions

An algorithm for nonlinear feature extraction has been presented. This algorithm emphasizes structures from graph theory such as the minimal spanning tree, inconsistent edges, and diameter edges as important attributes to be preserved under transformation. The process is sequential and hierarchical in nature thus easing the computational complexity encountered with other nonlinear feature extraction algortihms. Since the algorithm was formulated in such a manner that stressed structural properties present in data sets, the incorporation of this method in an interactive pattern recognition system would be interesting.

Encouraging results from the application of this algorithm to various data sets were presented. Currently, application of this method to other data bases is being performed. The reults from this effort should appear shortly.

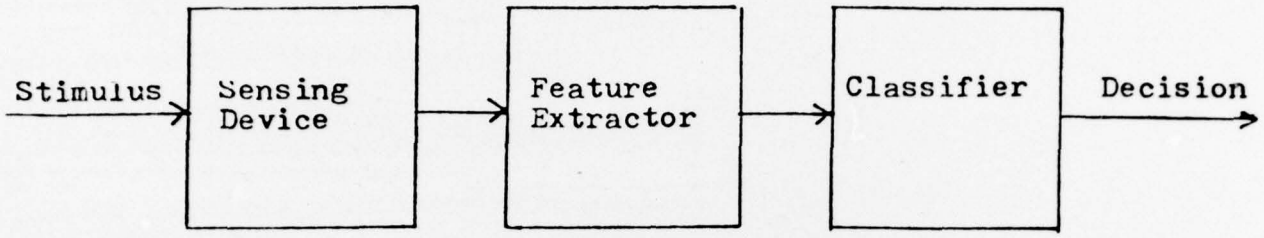

Fig. 1: Typical Pattern Recognition System

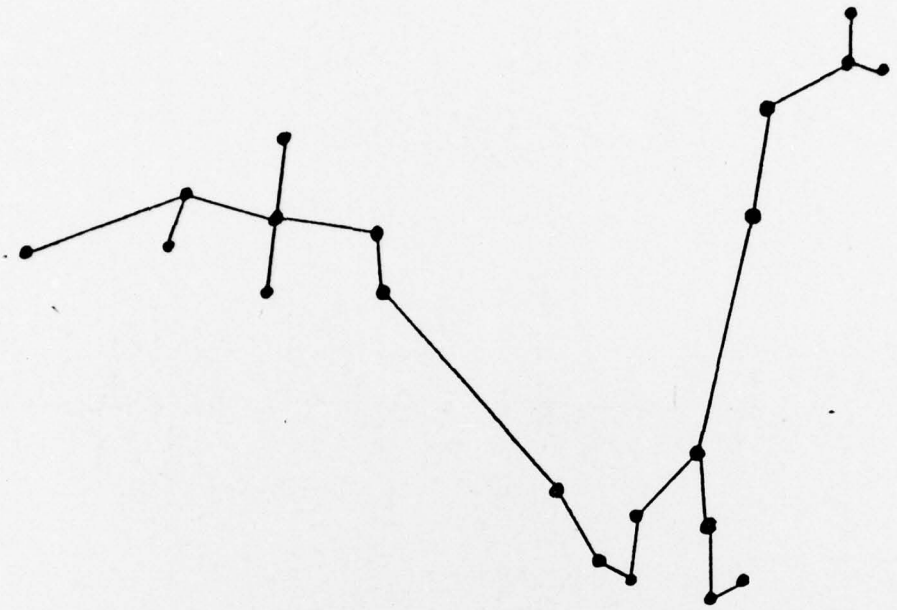

Fig. 2: Minimal Spanning Tree for Points in R^2

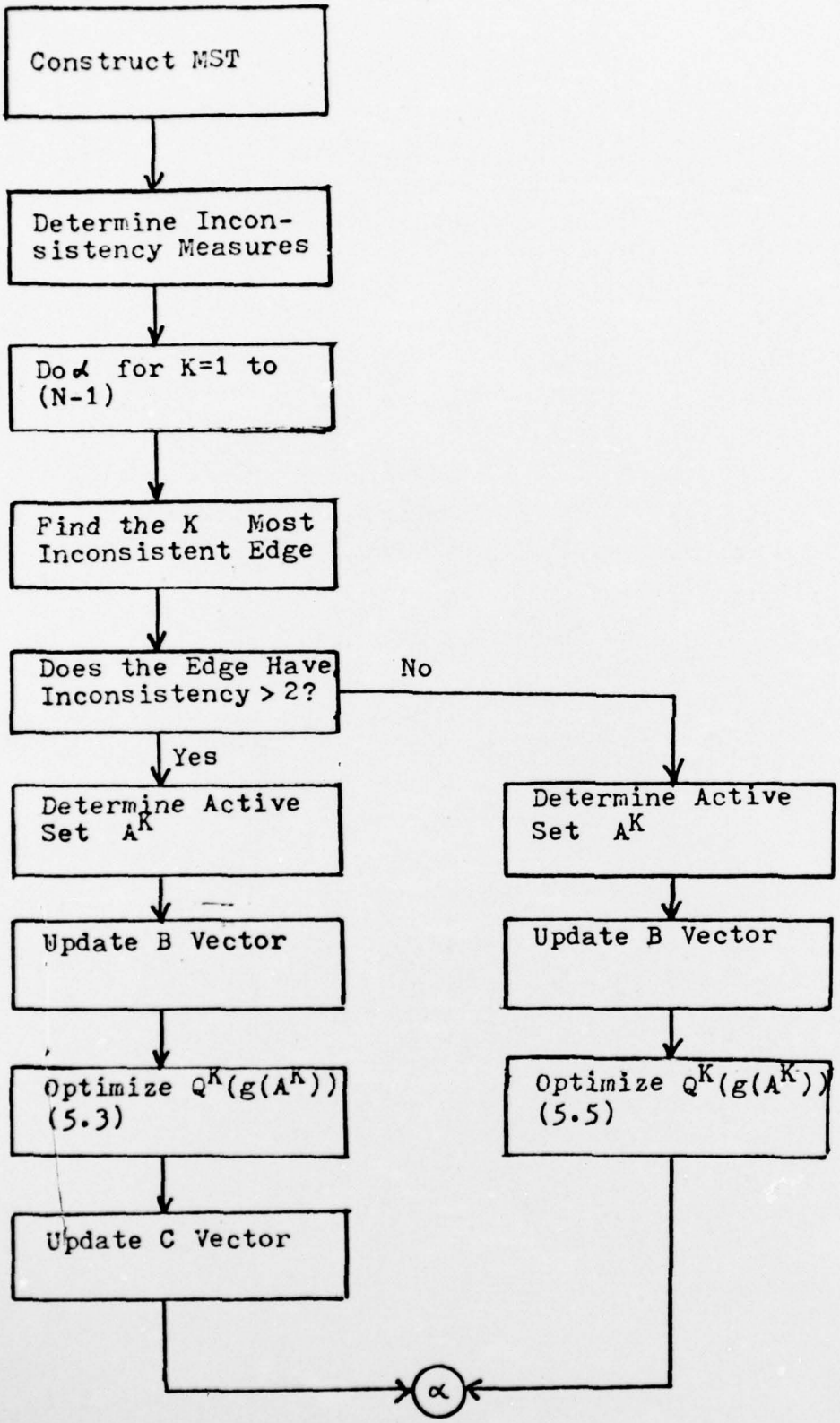

Fig. 3: Flowchart for Feature Extraction Algorithm

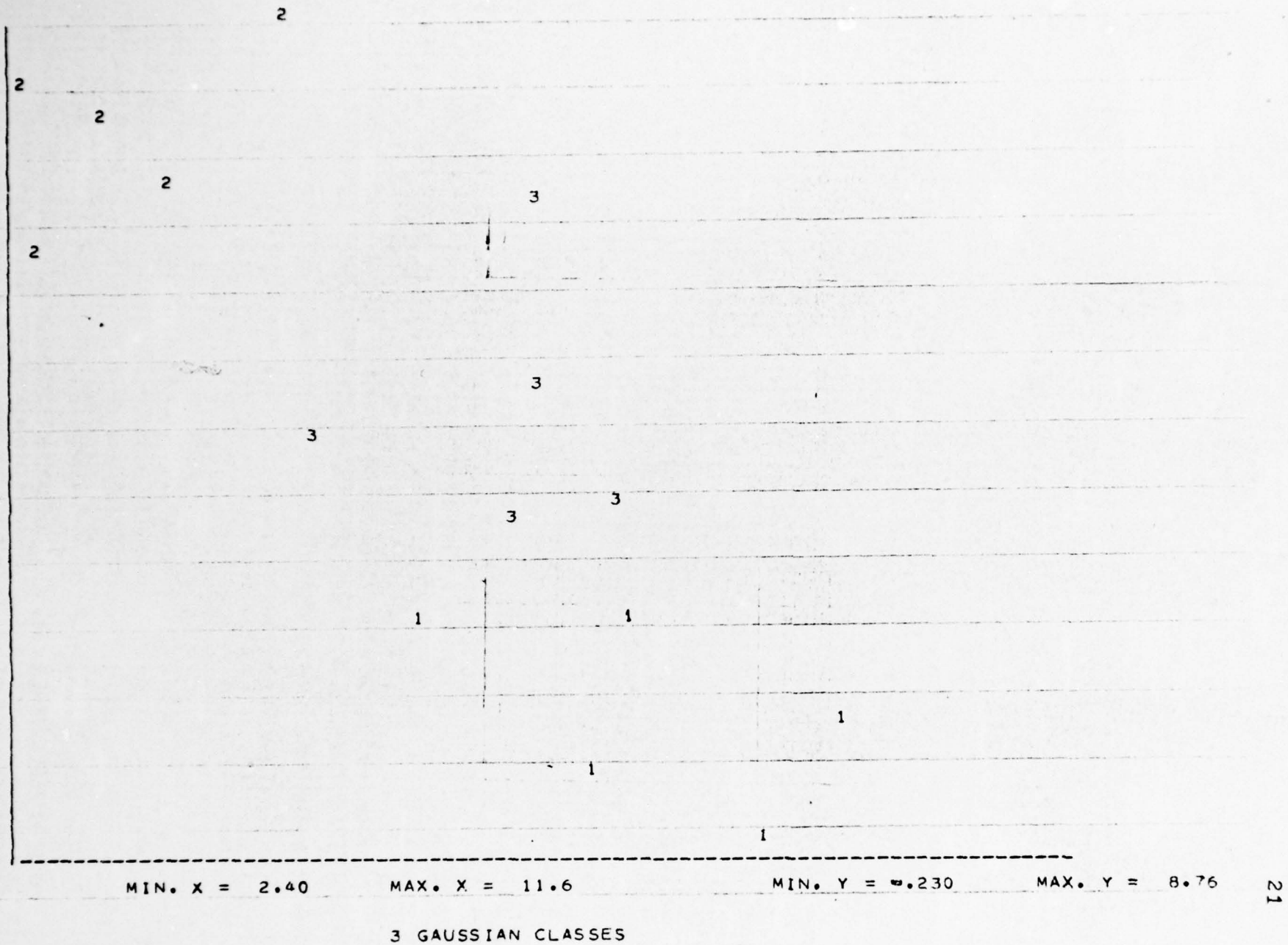

Fig. 4: Nonlinear Mapping for Artificial Data

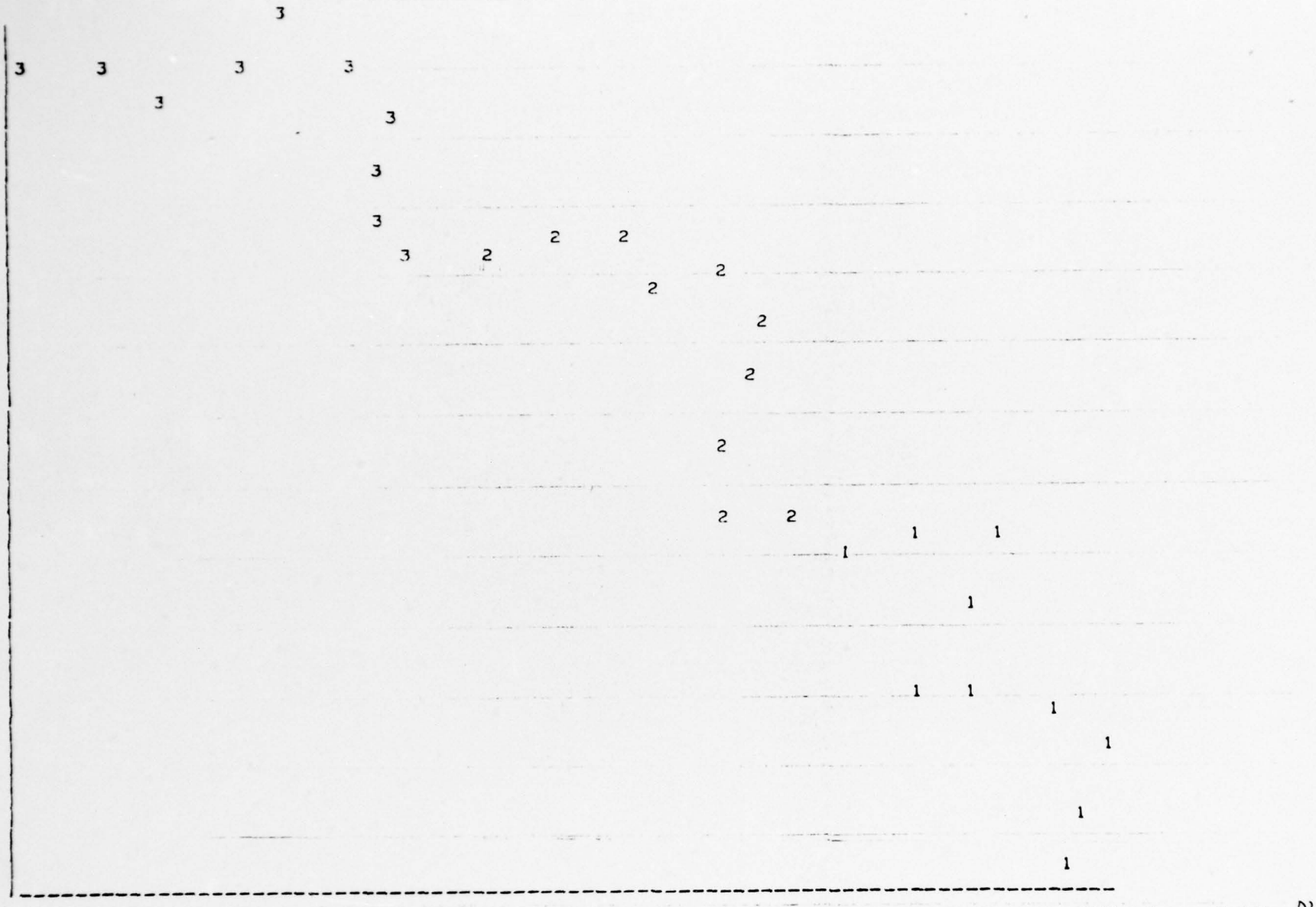

Fig. 5: Nonlinear Mapping for Nonlinear Data

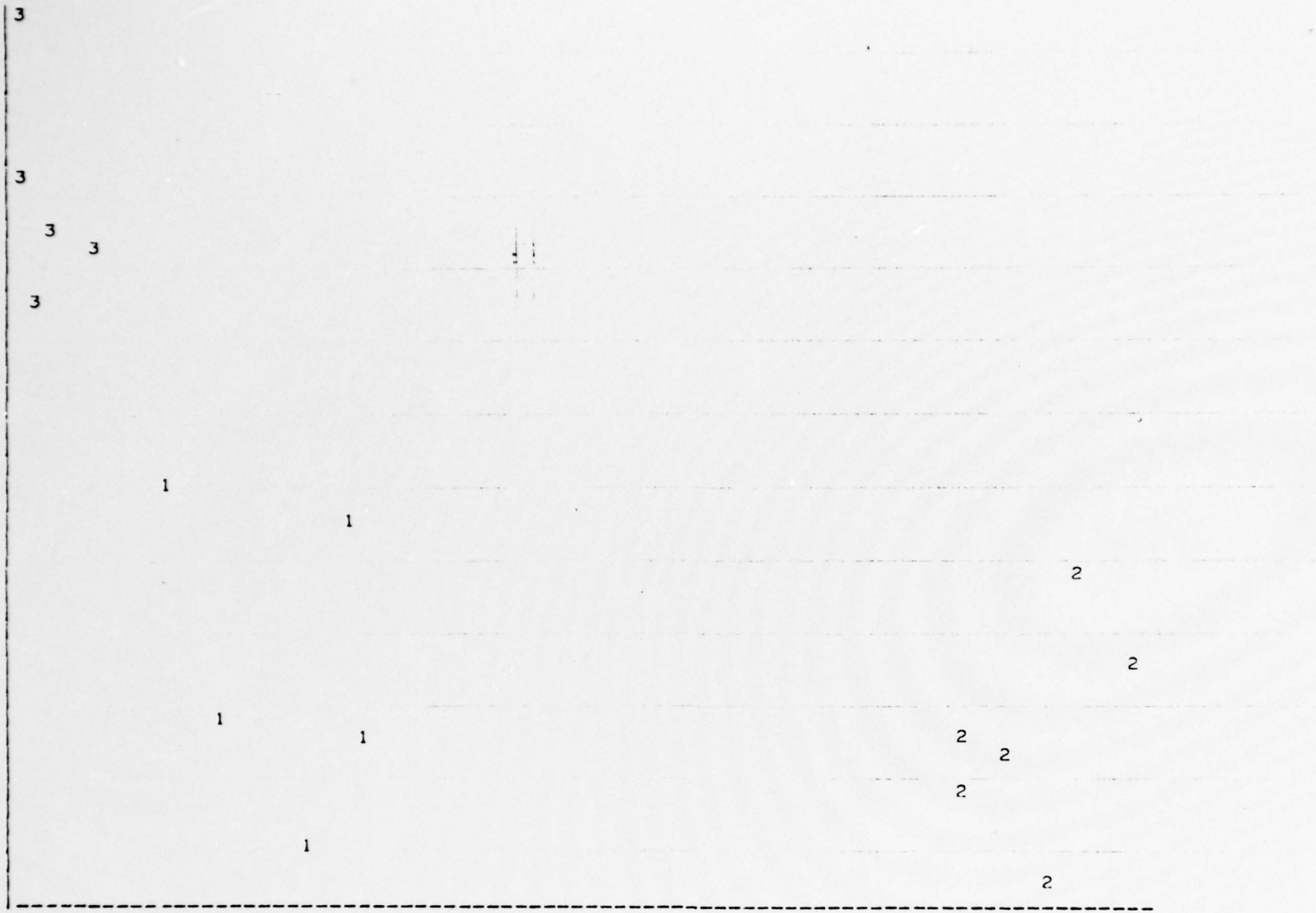

Fig. 6. Nonlinear Mapping for Oil Spectragraphic Data

REFERENCES

(1) Starks, S.A. and de Figueiredo, "A New Approach to Structure Preserving Feature Extraction", Proceedings of the 1977 Conference on Information Sciences and Systems, Baltimore, Maryland, March, 1977.

(2) Meisel, W.S., Computer-Oriented Approaches to Pattern Recognition, New York: Academic Press, 1972.

(3) Tou, J.T. and Gonzalez, R.C., Pattern Recognition Principles, Reading, Mass.: Addison-Wesley, 1974.

(4) Fu, K.S., Sequential Methods in Patten Recognition and Machine Learning, New York: Academic Press, 1968.

(5) Sammon, J.W., "A Nonlinear Mapping for Data Structure Analysis", IEEE Transactions on Computers, vol.C-18, no.5, pp.401-409, May 1969.

(6) Chang, C.L. and Lee, R.C.T., "A Heuristic Relaxation Method for Nonlinear Mapping in Cluster Analysis", IEEE Transactions on Systems, Man, and Cybernetics, March 1973. pp. 197-200.

(7) Shepherd, R.N., "Analysis of Proximities: Multidimensional Scaling with an Unknown Distance Function", Psychometrika, vol. 27, no. 2, June 1962.

(8) Calvert, T.W., "Nonorthogonal Projections for Feature Extraction in Pattern Recognition", IEEE Transactions on Computers, May 1970, pp. 447-452.

(9) White, I., "Comment on 'A Nonlinear Mapping for Data Structure Analysis'", IEEE Transactions on Computers, Feb. 1972, pp. 220-221.

(10) Duda, R.O. and Hart, P.E., Pattern Classification and Scene Analysis, New York: Wiley, 1973.

(11) Anderberg, M.R., Cluster Analysis for Application, New York: Academic Press, 1973.

(12) Hartigan, J.A., Clustering Algorithms, New York: Wiley, 1975.

(13) Johnson, S.C., "Hierarchical Clustering Schemes", Psychometrika, vol. 32, no. 3, pp. 241-254, 1967.

(14) Ore, O., Theory of Graphs, American Mathematical Colloquium Publications, vol. 38, 1962.

(15) Prim, R.C., "Shortest Connection Networks and Some Generalizations", Bell System Technical Journal, Nov. 1957, pp. 1389-1401.

(16) Zahn, C.T., "Graph-Theoretical Methods for Detecting and Describing Gestalt Clusters", IEEE Transactions on Computers, vol. C-20, no. 1, pp. 68-86, Jan. 1971.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AFOSR-TR-77-1141
2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER: EE-TR-7708

4. TITLE (and Subtitle): AN ALGORITHM FOR OPTIMAL NONLINEAR STRUCTURE PRESERVING FEATURE EXTRACTION.

5. TYPE OF REPORT & PERIOD COVERED: Interim rept.

6. PERFORMING ORG. REPORT NUMBER: EE 7708

7. AUTHOR(s): S. A. Starks and R. J. P. de Figueiredo

8. CONTRACT OR GRANT NUMBER(s): AFOSR 75-2777-754

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Rice University
Department of Electrical Engineering
Houston, TX. 77001

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
61102F
2304/A2

11. CONTROLLING OFFICE NAME AND ADDRESS
Air Force Office of Scientific Research (NM)
Bolling AFB, Washington DC 20332

12. REPORT DATE: June 1977

13. NUMBER OF PAGES: 26

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)

15. SECURITY CLASS. (of this report): UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

For presentation at the 1977 IEEE International Symposium on Information Theory to be held at Cornell University, Ithaca, N.Y. on October 10 through 14, 1977.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Feature extraction; pattern recognition; cluster analysis; data structures; nonlinear transformations; dimensionality reduction; data compression.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This paper presents a new approach to nonlinear structure preserving feature extraction. The ideas behind this method were first introduced in an earlier paper by the authors ("A New Approach to Structure Preserving Feature Extraction," in the Proc. of the 1977 Conference on Information Sciences and Systems, pp. 494-500, the Johns Hopkins University, Dept. of Electrical Engineering, Baltimore, Md., April 1977; also, Rice University Technical Report

(Continued on Reverse Side)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

463244

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ABSTRACT - Continued

EE 7707). This method is based on certain graph theoretical considerations (such as the minimal spanning tree, edge inconsistency, and diameter edges) and topological considerations (such as interpoint distance and measures). After introduction of the subject matter and appropriate background material, the algorithm is formulated, ~~in Section 5~~.

Numerical results from the application of this algorithm to various test data sets are presented. Evaluation of these test results are quite encouraging.